Le.31

VISUALIZATIONS OF MACHINE LEARNING WITH UNTARGETED AND TARGETED FEATURES FROM COMPREHENSIVE TWO-DIMENSIONAL GAS CHROMATOGRAPHY

Stephen E. Reichenbach[1,2], Claudia A. Zini[3], Juliane E. Welke[3], Chiara Cordero[4], Qingping Tao[2]

[1] Computer Science & Engineering Department - University of Nebraska, 260 Avery Hall, 68588-0115 Lincoln, United States

[2] GC Image, LLC - GC Image, LLC, PO Box 57403, 68505 Lincoln, United States

[3] Universidade Federal do Rio Grande do Sul, Institute of Chemistry, 91501-970 Porto Alegre, Brazil

[4] Università degli Studi di Torino, Dipartimento di Scienza e Tecnologia del Farmaco, I-10125 Turin, Italy

New visualization methods and tools for machine learning (ML) support both quick reviews of the overall consistency of individual samples to ML models as well as detailed reviews of the roles of individual chemical features in the sample-to-model relationships. These methods and tools are designed and well-suited for the high chemical-dimensionality in data produced by comprehensive two-dimensional gas chromatography (GCxGC) and other comprehensive multidimensional techniques. Advanced methods for automated untargeted and targeted (UT) feature analysis with template-based pattern recognition provide unambiguous correspondences between hundreds or even thousands of peak features across large numbers of multidimensional chromatograms. When general machine learning methods, such as Linear Discriminant Analysis (LDA) and K-Nearest-Neighbors, are applied to sets of high-dimensional chemical-feature vectors from multiple samples, it can be challenging to interpret the high-dimensional relationships of individual samples to models and the roles of individual chemical features in those relationships. The new visualization methods and tools plot sample features relative to the different models inherent in different ML methods. Customizing visualizations for different ML methods supports easier recognition of sample-to-model fits and the roles of individual features for individual compounds. The visualizations are demonstrated for machine learning of the characteristics of wine samples analyzed by GCxGC with mass spectrometry (GCxGC-MS).